# PATENT ABSTRACTS OF JAPAN

(11)Publication number : 2000-105768
(43)Date of publication of application : 11.04.2000

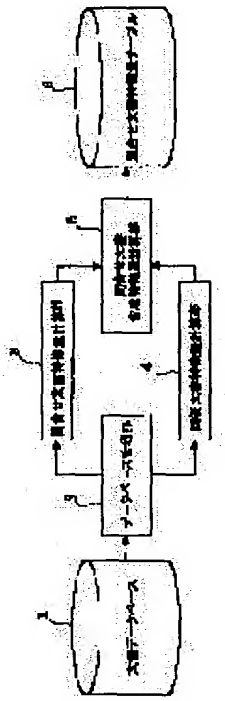
(51)Int.CI. G06F 17/30
G06F 17/27

(21)Application number : 10-273492
(22)Date of filing : 28.09.1998

(71)Applicant : NIPPON TELEGR & TELEPH CORP <NTT>
(72)Inventor : MORI DAIJIRO
OKUBO MASAKATSU
SUGIZAKI MASAYUKI
TANAKA KAZUO

## (54) DEVICE FOR CALCULATING FEATURE AMOUNT OF INQUIRY DOCUMENT, AND METHOD THEREFOR

(57)Abstract:
PROBLEM TO BE SOLVED: To provide a method for calculating the feature amount of an inquiry document for calculating feature amount suited to an inquiry content.
SOLUTION: An inquiry document feature amount calculating part 3 calculates inquiry document feature amount for quantitatively describing the features of an inquiry document for each inquiry document in an inquiry document group. An answer document feature amount calculating part 4 calculates answer document feature amount for quantitatively describing the features of each answer document for each answer document in an answer document group. An inquiry document synthetic feature amount calculating part 5 inputs the inquiry document feature amount and the answer document feature amount, calculates correction feature amount corresponding to the answer document feature amount according to a prescribed relation, synthesizes the inquiry document feature amount with the correction feature amount, and outputs the synthesizsed result as the final feature amount of the inquiry document.



LEGAL STATUS

[Date of request for examination]

[Date of sending the examiner's decision of rejection]

[Kind of final disposal of application other than the examiner's decision of rejection or application converted registration]

[Date of final disposal for application]

[Patent number]

[Date of registration]

[Number of appeal against examiner's decision of rejection]

[Date of requesting appeal against examiner's decision of rejection]

[Date of extinction of right]

(19)日本国特許庁（ＪＰ）　(12)公開特許公報（A）　(11)特許出願公開番号
特開2000－105768
（P2000－105768A）
(43)公開日　平成12年４月11日(2000.4.11)

| (51)Int.Cl.' | 識別記号 | FI | | テーマコード'(参考) |
|---|---|---|---|---|
| G06F 17/30 | | G06F 15/401 | 320A | 5B075 |
| 17/27 | | 15/38 | D | 5B091 |
| | | 15/403 | 350C | |

審査請求　未請求　請求項の数８　ＯＬ　（全 7 頁）

(21)出願番号　特願平10－273492

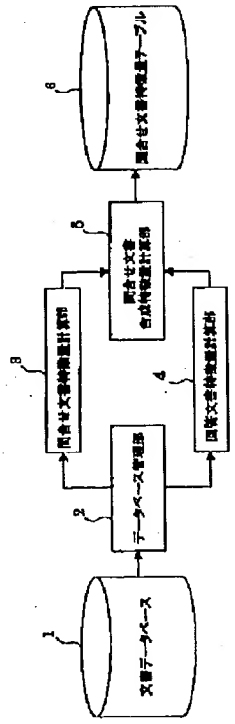
(22)出願日　平成10年９月28日(1998.9.28)

(71)出願人　000004226
日本電信電話株式会社
東京都千代田区大手町二丁目３番１号
(72)発明者　森　大二郎
東京都新宿区西新宿三丁目19番２号　日本電信電話株式会社内
(72)発明者　大久保　雅且
東京都新宿区西新宿三丁目19番２号　日本電信電話株式会社内
(74)代理人　100070219
弁理士　若林　忠　（外２名）

最終頁に続く

(54)【発明の名称】　問合わせ文書の特徴量計算装置および方法

(57)【要約】
【課題】　問合せの内容に適合した特徴量を計算することができる、問合わせ文書特徴量の計算方法およびその装置を提供する。
【解決手段】　問合せ文書特徴量計算部３は、問合せ文書の特徴を定量的に記述する問合せ文書特徴量を問合せ文書集合中の各問合せ文書について計算する。回答文書特徴量計算部４は、各回答文書の特徴を定量的に記述する回答文書特徴量を回答文書集合中の各回答文書について計算する。問合せ文書合成特徴量計算部５は、問合せ文書特徴量と回答文書特徴量とを入力し、所定の関係によって回答文書特徴量に対応する修正特徴量を演算し、問合せ文書特徴量と修正特徴量とを合成してその合成結果を当該問合せ文書の最終特徴量として出力する。



1

【特許請求の範囲】

【請求項１】　複数の問合せ文書と、各問合せに対する回答文書とを対応付けて管理し、各問合せ文書の特徴量を計算する特徴量計算装置において、

問合せ文書の特徴を定量的に記述する問合せ文書特徴量を、問合せ文書集合中の各問合せ文書について計算する、問合せ文書特徴量計算手段と、

各回答文書の特徴を定量的に記述する回答文書特徴量を、回答文書集合中の各回答文書について計算する、回答文書特徴量計算手段と、

前記問合せ文書特徴量と前記回答文書特徴量とを入力し、所定の関係によって前記回答文書特徴量に対応する修正特徴量を演算し、問合せ文書特徴量と修正特徴量とを合成してその合成結果を当該問合せ文書の最終特徴量として出力する問合せ文書合成特徴量計算手段と、を備えていることを特徴とする問合せ文書特徴量計算装置。

【請求項２】　問合せ文書合成特徴量計算手段は、回答文書特徴量に所定の定数を乗算し、その乗算結果を修正特徴量として、当該回答文書に対応する問合せ文書の最終特徴量を生成する手段を有する、請求項１に記載の問合せ文書特徴量計算装置。

【請求項３】　問合せ文書合成特徴量計算手段は、問合せ文書の特徴と、当該問合せ文書に対応する回答文書の特徴との間の共通する特徴成分を抽出して定量的に評価する手段を有し、その共通する特徴成分の評価量を修正特徴量として当該問合せ文書の最終特徴量を生成する手段を有する、請求項１に記載の問合せ文書特徴量計算装置。

【請求項４】　問合せ文書合成特徴量計算手段は、最終特徴量を求めようとする問合せ文書$q_i$に対応する回答文書$a_i$と、回答文書集合中の回答文書$a_s$との間の類似度$R(i,s)$を演算する手段を有し、前記回答文書$a_s$に対応する問合せ文書$q_s$の問合せ文書特徴量$FV(q_s)$に類似度$R(i,s)$を乗算してその乗算結果$R(i,s)FV(q_s)$を問合せ文書集合中の、当該問合せ文書$q_i$以外の総ての問合せ文書について合成し、その合成結果を修正特徴量として、当該問合せ文書$q_i$の問合せ文書特徴量$FV(q_i)$と合成して当該問合せ文書$q_i$の最終特徴量を生成する請求項１に記載の問合せ文書特徴量計算装置。

【請求項５】　複数の問合せ文書と、各問合せに対する回答文書とを対応付けて管理し、各問合せ文書の特徴量を計算する問合せ文書特徴量計算方法において、

問合せ文書の特徴を定量的に記述する問合せ文書特徴量を、問合せ文書集合中の各問合せ文書について計算する問合せ文書特徴量計算処理と、

回答文書の特徴を定量的に記述する回答文書特徴量を、回答文書集合中の各回答文書について計算する回答文書特徴量計算処理と、

前記問合せ文書特徴量と前記回答文書特徴量とに基づいて、所定の関係によって前記回答文書特徴量に対応する修正特徴量を演算し、問合せ文書特徴量と修正特徴量とを合成してその合成結果を当該問合せ文書の最終特徴量として出力する問合せ文書合成特徴量計算処理を含んでいることを特徴とする問合せ文書特徴量計算方法。

2

【請求項６】　問合せ文書合成特徴量計算処理は、回答文書特徴量に所定の定数を乗算し、その乗算結果を修正特徴量として、当該回答文書に対応する問合せ文書の最終特徴量を生成する過程を有する、請求項５に記載の問合せ文書特徴量計算方法。

【請求項７】　問合せ文書合成特徴量計算処理は、問合せ文書の特徴と回答文書の特徴との間の共通する特徴成分を抽出して定量的に評価する過程を有し、その共通する特徴成分の評価量を修正特徴量として当該問合せ文書の最終特徴量を生成する過程を有する、請求項５に記載の問合せ文書特徴量計算方法。

【請求項８】　問合せ文書合成特徴量計算処理は、最終特徴量を求めようとする問合せ文書$q_i$に対応する回答文書$a_i$と、回答文書集合中の回答文書$a_s$との間の類似度$R(i,s)$を演算する過程を有し、前記回答文書$a_s$に対応する問合せ文書$q_s$の問合せ文書特徴量$FV(q_s)$に類似度$R(i,s)$を乗算してその乗算結果$R(i,s)FV(q_s)$を問合せ文書集合中の、当該問合せ文書$q_i$以外の総ての問合せ文書について合成し、その合成結果を修正特徴量として、当該問合せ文書$q_i$の問合せ文書特徴量$FV(q_i)$と合成して当該問合せ文書$q_i$の最終特徴量を生成する、請求項5に記載の問合せ文書特徴量計算方法。

【発明の詳細な説明】

【０００１】

【発明の属する技術分野】本発明は、問合せに対して回答を行う業務を通して得られる情報を蓄積し、検索・分類・自動応答などの諸機能を実現する際に必要な問合せ文書の特徴量計算装置および方法に関する。

【０００２】

【従来の技術】計算機技術の発展に伴い、大量の蓄積文書を対象として検索や分類を行うことが可能になった。検索や分類などの処理を高速かつ高精度に行うために、対象となる文書の特徴量をあらかじめ抽出する技術が一般に用いられる。文書の特徴量を計算する方法として、以下のものが知られている。

【０００３】まず文書を、文字列や単語や文節を単位とする要素に分解し、要素の出現頻度や長さによってその重要度を算出する。これらの要素毎の重要度を成分とするベクトル（各要素を基底とし、重要度を成分とするベクトル）を文書の特徴量とする。

【０００４】あるいは、単語などの要素を直接ベクトル空間の基底にとるのではなく、文書に含まれる要素間の関連度を所定の方法で算出し、関連度の高い要素が相互に近傍に位置するように、予め定められたｎ次元のベクトル空間上に要素を適宜配置して各要素をそのｎ次元のベクトル空間上のベクトルに対応させておき、文書の特

徴を計算するとき当該文書を構成する各要素を抽出し、抽出された要素に対応するベクトルの、上記ベクトル空間上におけるベクトル和を求めて特徴量を計算する方法も知られている。

【０００５】

【発明が解決しようとする課題】上記の従来の技術においては、いずれも、特徴を求める対象となる問合せ文書そのものに含まれる要素を抽出し、これに基づいて文書の特徴量を計算している。

【０００６】しかし、不特定多数の人から受ける問合せ文書は、使用される語彙や表現が人によって異なる傾向が強い。したがって、文書に含まれる要素から特徴量を計算すると、問合せの内容が同一であっても、特徴量が一致しない場合が多く発生する。

【０００７】本発明は、上述の従来の技術に見られる課題に鑑みてなされたもので、従来の方法よりも問合せの内容に適合した特徴量を計算することができる、問合わせ文書の特徴量計算方法およびその装置を提供することを目的とする。

【０００８】

【課題を解決するための手段】従来の技術においては、問合せ文書から抽出される特徴のみによって特徴量を計算していたのに対して、本発明では、該問合せ文書に対応する回答文書から抽出される特徴に基づいて問合せ文書の特徴量を修正することによって、最終的な特徴量を計算する。

【０００９】そのために、本発明の問合せ文書特徴量計算装置は、問合せ文書特徴量計算部と回答文書特徴量計算部と問合せ文書合成特徴量計算部とを備えている。問合せ文書特徴量計算部は、問合せ文書の特徴を定量的に記述する問合せ文書特徴量を問合せ文書集合中の各問合せ文書について計算する。回答文書特徴量計算部は、各回答文書の特徴を定量的に記述する回答文書特徴量を回答文書集合中の各回答文書について計算する。問合せ文書合成特徴量計算部は問合せ文書特徴量と回答文書特徴量とを入力し、所定の関係によって回答文書特徴量に対応する修正特徴量を演算し、問合せ文書特徴量と修正特徴量とを合成してその合成結果を当該問合せ文書の最終特徴量として出力する。

【００１０】本発明の問合せ文書特徴量計算方法は、複数の問合せ文書と、各問合せに対する回答文書とを対応付けて管理し、各問合せ文書の特徴量を計算する問合せ文書特徴量計算方法であって、問合せ文書の特徴を定量的に記述する問合せ文書特徴量を、問合せ文書集合中の各問合せ文書について計算する問合せ文書特徴量計算処理と、回答文書の特徴を定量的に記述する回答文書特徴量を、回答文書集合中の各回答文書について計算する回答文書特徴量計算処理と、問合せ文書特徴量と前記回答文書特徴量とに基づいて、所定の関係によって前記回答文書特徴量に対応する修正特徴量を演算し、問合せ文書特徴量と修正特徴量とを合成してその合成結果を当該問合せ文書の最終特徴量として出力する問合せ文書合成特徴量計算処理を含んでいる。



【００１１】

【作用】前述のように、問合せ文書において、使用される語彙や表現が多様であるため、そこから抽出される特徴量が、問合せの内容と合致しない場合がある。一方、不特定多数の人から受ける問合せに回答する業務においては、問合せ文書を作成する人の数に比べて、回答文書を作成する人の数の方が少ないので、回答文書に現れる語彙や表現は、問合せ文書のそれと比べてより一様であり、同一の問合せ内容に対しては、同一の語彙や表現を用いた回答文書が作成される傾向が強い。したがって、問合せ文書から抽出された特徴量に対して、所定の関係によって回答文書特徴量に対応する修正特徴量を演算し、問合せ文書特徴量と修正特徴量とを合成することによって、従来の技術よりも問合せ内容に適合した最終特徴量を計算することが可能になる。

【００１２】修正特徴量としては、回答文書特徴量に所定の定数を乗算し、その乗算結果を修正特徴量とすることができる。また、問合せ文書の特徴と回答文書の特徴との間の共通する特徴成分を抽出して定量的に評価し、その共通する特徴成分の評価量を修正特徴量とすることができる。さらに、回答文書集合中の２つの回答文書の組の類似度を計算し、その類似度を重率として問合せ文書特徴量を合成し、その合成結果を修正特徴量とすることもできる。このようにして、回答文書から抽出される特徴を問い合わせ文書特徴量に反映させることができる。

【００１３】

【発明の実施の形態】次に、図面を参照して本発明の特徴量計算装置の実現形態を説明する。本実施形態の特徴量計算装置は、本発明の特徴量計算方法を実施するための装置である。

【００１４】図１は特徴量計算装置のシステム構成を示すブロック図である。図２は、図１のデータベース管理部２に保持されている問合せ・回答対応表の概念図である。本実施形態の特徴量計算装置は、文書データベース１、データベース管理部２、問合せ文書特徴量計算部３、回答文書特徴量計算部４、問合せ文書合成特徴量計算部５、問合せ文書特徴量計算部６を備えている。

【００１５】文書データベース１には、問合せ文書と回答文書が格納されている。データベース管理部２は、文書データベース１から問合せ文書または回答文書を取り出す機能を有する。さらに、データベース管理部２は、図２に示すような問合せ・回答対応表を備えており、問合せ文書をキーとして該問合せ文書に対応する回答文書を取り出し、また、回答文書をキーとして、回答文書に対応する問合せ文書を取り出すことができる。図２の向かって右側の図は図１の文書データベース１であり、図

２の向かって左側の図が問合せ・回答対応表を表す。図示されているように、問合せ・回答対応表はレコード番号フィールド、問合せ文書識別情報フィールド、当該問合せ文書に対応する回答文書の回答文書識別情報フィールドを含んでいる。

【００１６】回答文書特徴量計算部４および問合せ文書特徴量計算部３は、複数の成分から構成される特徴量を、文書の内容に即して抽出する。これらの特徴量計算部３、４は特徴を構成する複数の成分（以下、特徴成分と記す）から構成される特徴量を、文書の内容に即して抽出する手段であれば、いずれも本発明において適用可能である。本実施形態では、文書の各文を形態素解析して単語に分解し、各単語の出現頻度を成分とする特徴ベクトルによって特徴量を表すものとする。

【００１７】問合せ文書合成特徴量計算部５は、問合せ文書特徴量と、回答文書特徴量を合成し、最終的な問合せ文書特徴量を生成し、問合せ文書特徴量テーブル６に、問合せ文書番号と最終的な問合せ文書特徴量を記録する。

【００１８】図１の問合せ文書特徴量計算装置は次のように動作する。

【００１９】データベース管理部２は文書データベース１から、対応する問合せ文書と回答文書を取り出し、それぞれについて問合せ文書特徴量計算部および回答文書特徴計算部により問合せ文書特徴量および回答文書特徴量を計算する。問合せ文書合成特徴量計算部５は、問合せ文書特徴量と、所定の関係で回答文書特徴量に対応する修正特徴量とを合成し、問合せ文書特徴量テーブル６に、問合せ文書番号と最終的な問合せ文書特徴量（以



下、最終問合せ文書特徴量と記す）を記録する。次に、問合せ文書特徴量計算部３および回答文書特徴計算部４（以下、特徴量計算部と総称する）、問合せ文書合成特徴量計算部５の動作について更に詳細に説明する。

【００２０】図３は特徴量計算部の処理フローを示す図である。まず、各文書ｉ（１≦ｉ≦M、Mは文書総数）を形態素解析し（ステップＳ１）、文書の各文を単語に分解する。次に、単語リストを生成し（ステップＳ２）、各単語ｊ（１≦ｊ≦N、Nは文書集合における全単語数）の出現頻度tf(i, j)を計算する（ステップＳ３）。次に、各単語ｊ毎に重み付け処理をする（ステップＳ４）。本実施形態においては、重率は、log(M/df(j))とする。ここで、df(j)は文書集合における単語ｊの出現回数である。（この重率は次の意味をもつ。一般にある文書に「特定」の単語が高い頻度で使用されているときには、その文書は、その特定単語の内容によって特徴付けられる。しかし、ある単語が、どの文書にも共通して高い頻度で使用されている場合には、その単語は当該文書を特徴付ける単語ということはできない。df(j)/Mは、単語jの１文書当たりの出現頻度である。df(j)/Mが大きいということは、その単語が、どの文書にも共通して高い頻度で使用されていることを意味する。log(M/df(j))は、１文書当たり、平均１０回の出現頻度の単語に、１文書当たり、平均１００回の出現頻度をもつ単語の２倍の重率を与える重み付け処理である）。

【００２１】この重み付け処理によって文書ｉの特徴ベクトルFV(i)のｊ成分（単語1, 2, 3.. j..Nを基底とするN次元ベクトル空間のｊ成分）w(i, j)は、

$$w(i,j) = tf(i,j) \cdot \log(M/df(j)) \qquad (1)$$

になる。したがって、文書ｉの特徴ベクトルFV(i)は次式で表される。

【００２２】

$$FV(i) = (w(i,1), \ldots\ w(i,j), \ldots\ w(i,N)) \qquad (2)$$

このようにして、特徴量計算部３，４は文書に含まれる各単語に基づいて、特徴ベクトルFV(i)を計算する。

【００２３】ここでは、ベクトルの成分は単語全般としているが、文書の特徴を表すのに適当な単位として、自立語のみを要素とする方法や、接辞や複合語を含めて要素とする方法、名詞句に含まれる単語列を要素とする方法をも用いることができる。また、問い合わせ文書特徴量計算部３と解答文書特徴量計算部４とで、異なる方法で特徴量を計算することができる。

【００２４】図４は問合せ文書集合と回答文書集合との対応を示す図である。文書集合は問合せ文書$q_i$と、対応する回答文書$a_i$との対で構成されている。しかし、本実施形態においては、式(1)、式(2)のような数値を計算するときには、問合せ文書集合と回答文書集合は独立の文書集合と見做し、前記の各数値も独立に算出するものとする。また、問合せ文書と、これに対応する回答文書については文書番号は同一とする。

【００２５】次に、本実施形態の問合せ文書合成特徴量計算部５について説明する。本実施形態の問合せ文書合成特徴量計算部５は、回答文書特徴量と問合せ文書特徴量との両者に基づいて最終問合せ文書特徴量を計算する。

【００２６】問合せ文書合成特徴量計算部５の第１の実施例においては、回答文書の特徴ベクトル

$$FV(a_i) = (a(i,1),\ a(i,2)\ldots,\ a(i,j)\ldots,\ a(i,S)) \qquad (3)$$

に所定の重率Cを乗算して問合せ文書の特徴ベクトル

$$FV(q_i) = (q(i,1),\ q(i,2)\ldots,\ q(i,j)\ldots,\ q(i,S)) \qquad (4)$$

に加算して、その加算結果、すなわち、

$$FV'(q_i) = (q(i,1)+C\ a(i,1),\ldots,q(i,j)+C\ a(i,j),\ldots,\ q(i,S)+C\ a(i,S))$$

を問合せ文書の最終特徴ベクトルとする。ここで、$q_i$、$a_i$はそれぞれ文書番号ｉの問合せ文書および回答文書である。また、a(i,j)は文書番号ｉの回答文書の特徴ベクトルFV($a_i$)のj成分（該成分が存在しなければ０）である。q(i,j)は文書番号ｉの問合せ文書の特徴ベクトルFV($q_i$)のj成分（該成分が存在しなければ０）である。Ｓは回答文書および問合せ文書を合わせた全文書集合における全単語数である。本実施例では、式(5)のC a(i,j)が修正特徴量のj成分である。

【００２７】問合せ文書合成特徴量計算部５の第２の実施例においては、問合せ文書の特徴ベクトルと回答文書

$$FV'(q_i) = (q(i,1)\cdot(1+a(i,1)),\ q(i,2)\cdot(1+a(i,2)),\ldots,\ q(i,j)\cdot(1+a(i,j)),\ldots,\ q(i,N)\cdot(1+a(i,N)),) \quad (6)$$

ここで、q(i,j)は、問合せ文書ｉの特徴ベクトルの単語ｊに対応する成分、a(i,j)は、回答文書ｉの特徴ベクトルの単語ｊに対応する成分（該成分が存在しなければ０）、Ｎは、問合せ文書集合における全単語数である。本実施例においては、q(i,k)a(i,k)が修正特徴量のk成分である。

【００２８】問合せ文書合成特徴量計算部５の第３の実施例において、問合せ文書$q_i$の最終特徴ベクトルFV'($q_i$)を計算しようとするとき、先ず、回答文書集合A中の総ての回答文書$a_s$(1≦s≦M)について、前掲の回答文書特徴量計算部３および４によって問合せ文書特徴ベクトルFV($q_s$)および回答文書特徴ベクトルFV($a_s$)（1≦s≦M)を計算する。ここで、Mは回答文書集合Aにおける回答文書集合の要素（回答文書）の総数である。次に、計算された回答文書特徴ベクトルFV($a_s$)（1≦s≦M)を用いて、当該問合せ文書$q_i$に対する回答文書$a_i$(∈A)と、回答文書集合A中の回答文書$a_s$(1≦s≦M)との総ての組合せについて類似度R(i,s)を計算する。類似度R(i,s)の計算方法は後述する。次に、回答文書$a_i$と組合せた回答文書$a_s$(1≦s≦M)に対応する問合せ文書$q_s$(1≦s≦M)の問合せ文書特徴ベクトルFV($q_s$)に類似度R(i,s)を重率として乗算し、その乗算結果を総てのsについてベクトル合成する。この合成結果を問合せ文書$q_i$の最終特徴ベクトルFV'($q_i$)とする。

【００２９】類似度R(i,s)は回答文書特徴ベクトルFV($a_i$)と 回答文書特徴ベクトルFV($a_s$)とのN次元空間（Nは単語要素の総数）における夾角に対応する。すなわち、

【００３０】

【数１】

(5)

の特徴ベクトルの間で共通する成分を抽出し、該共通部分を重みとして重み付けをすることによって最終的な問合せ文書特徴量を計算する。本実施例においては、問合せ文書（文書番号ｉ）の特徴ベクトルと回答文書（文書番号ｉ）の特徴ベクトルの間の共通する成分は各単語要素k（1≦k≦N)が問合せ文書と回答文書との両者に出現する頻度である。本実施例においては、この頻度をq(i,k)a(i,k)によって評価する（単語要素kがどちらか一方にしか含まれていなければ、q(i,k)a(i,k)は0になる）。本実施例における文書番号ｉの問合せ文書の最終特徴ベクトルFV'($q_i$)は次式で表される。

$$FV'(q_i) = \sum_{s=1}^{M} R(i,s)FV(q_s) \quad (7)$$

$$R(i,s) = \frac{\langle FV(a_i)FV(a_s)\rangle}{|FV(a_i)||FV(a_s)|} \quad (8)$$

$$\langle FV(a_i)FV(a_s)\rangle = \sum_{j=1}^{N} a(i,j)a(s,j) \quad (9)$$

$$|FV(a_i)| = \left(\sum_{j=1}^{N} a^2(i,j)\right)^{1/2} \quad (10)$$

$$|FV(a_s)| = \left(\sum_{j=1}^{N} a^2(s,j)\right)^{1/2} \quad (11)$$

式(7)～(11)において、式(10)および(11)はそれぞれ回答文書特徴ベクトルFV($a_i$)、FV($a_s$)の大きさ（成分の２乗和の平方根）を表す。式(8)、式(9)の記号<>はベクトルの内積（対応する成分の積和）を表す。したがって、式(8)のR(i,s)は回答文書特徴ベクトルFV($a_i$)とFV($a_s$)との夾角の余弦を表す。式(7)の加算範囲の上限Mは回答文書集合Aにおける回答文書の総数である（これは、問合せ文書集合における問合せ文書の総数に等しい（図４参照））。しかし、式(7)の和は類似度R(i,s)が所定の閾値以上の値をもつ項のみについて加算を実行して演算量を少なくすることができる。

【００３１】第３の実施例は次のような考え方に基いている。企業等のヘルプデスク等においては、不特定多数の顧客から寄せられる問合せに対して、特定の少数の対応要員から回答が返される。不特定多数の顧客から寄せられる問合せは、問合せの内容が同一であっても独自の語彙や言い回しが用いられることが多く、使用される語彙がまちまちである。しかし、特定の少数の対応要員から返される回答文書は、問合せの内容が同一であれば、高い類似性を示すことが期待できる。第３の実施例は、この仮定に基づき、回答文書間の類似度によって対応する問合せ文書の特徴ベクトルを調整したものである。

【００３２】一般に、２つの文書の特徴ベクトルを加算すると、両方の文書に共通に使用されている単語要素に対応する成分は加算されてその成分の値は増加するけれ

ど、共通に使用されていない単語要素に対応する成分の値は変わらない。したがって、２つの特徴ベクトルを加算して新たな特徴ベクトルを作ると、その新たな特徴ベクトルは、元の２つの特徴ベクトルに共通に含まれている単語要素に対応する成分が強調された特徴ベクトルになる。したがって、問合せ文書集合の中から、同一または類似の問合せ内容をもつ問合せ文書を選択して、それらの問合せ文書の特徴ベクトルを加算すると、それらの問合せ文書に共通に使用されている単語要素に対応する成分が強調された新たな特徴ベクトルが得られる。このとき、同一または類似の問合せ内容をもつ問合せ文書に共通に使用されている単語要素は、その問合せ内容の特徴を示す単語要素であると考えることができる。本実施例の類似度は、同一または類似の問合せ内容をもつ問合せ文書に大きな重率を与えることによって、同一または類似の問合せ内容をもつ問合せ文書が最終問合せ文書特徴ベクトルに対して大きな寄与を与えるように働く。この類似度を、問合せ文書の特徴ベクトルでなく、回答文書の特徴ベクトルによって計算することが本実施形態の特徴である。

【００３３】最後に、本発明は、特許請求の範囲に記載されている本発明の主旨を逸脱しない範囲でシステム構成や実現手段を変更することができる。例えば、文書の特徴ベクトルを記述するための単語要素として、自立語のみを単語要素とする方法や、接辞や複合語を含めて単語要素とする方法、名詞句に含まれる単語列を単語要素とする方法をも用いることができる。また、文書の特徴ベクトルの計算方法も、必ずしも式(1)、式(2)の方法である必要はない。大切なことは、任意の方法で計算された回答文書の特徴量によって、任意の方法で計算された問合せ文書の特徴量を修正するということが本発明の特許請求の範囲の趣旨である。また、第２の実施例において、単語要素k（1≦k≦N)が問合せ文書と回答文書との両者に出現する頻度を、必ずしもq(i,k)a(i,k)によって



評価する必要はない。大切なことは、問合せ文書と回答文書との両者に出現する単語要素の頻度によって、問合せ文書の特徴量を修正するということであって、その頻度の計算方法は、任意の方法を採用することができる。さらに、第３の実施例においては、類似度は、必ずしも式(7)～(11)の方法で計算する必要はない。大切なことは、問合せ文書の類似度ではなく、回答文書の類似度を用いることであって、その計算方法は任意の方法を採用することができる。

【００３４】

【発明の効果】以上説明したように、本発明においては、問合せ文書の特徴量を回答文書の特徴量によって修正することによって、問合せ文書が不特定多数の問合せ者による、独自の語彙や言い回しが用いられた問合せ文書であっても、問合せ文書の内容に適合した特徴量を計算することができる効果がある。その結果、問合せ文書の検索・分類・自動応答などの諸機能を従来より高い精度で実現することができるという効果がある。

【図面の簡単な説明】

【図１】本発明の特徴量計算装置のシステム構成を示すブロック図である。

【図２】図１のデータベース管理部２に保持されている問合せ・回答対応表の概念図である。

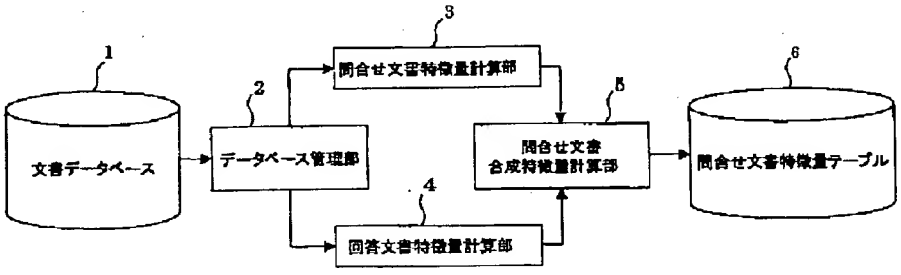
【図３】特徴量計算部の処理フローを示す図である。

【図４】問合せ文書集合と回答文書集合との対応を示す図である。

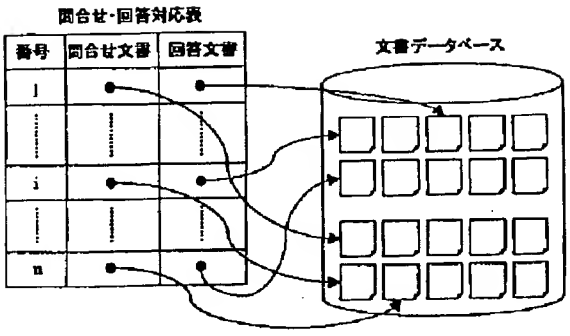
【符号の説明】

1　文書データベース
2　データベース管理部
3　問合せ文書特徴量計算部
4　回答文書特徴量計算部
5　問合せ文書合成特徴量計算部
6　問合せ文書特徴量テーブル

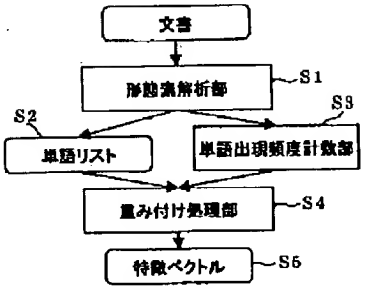
【図１】



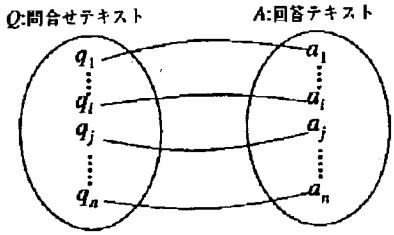
【図２】




【図３】




【図４】




---

フロントページの続き

(72)発明者　杉崎　正之
東京都新宿区西新宿三丁目19番２号　日本電信電話株式会社内

(72)発明者　田中　一男
東京都新宿区西新宿三丁目19番２号　日本電信電話株式会社内

Ｆターム(参考)　5B075 ND03 NR02 NS01 PR06 QM08
QP02 QS01 UU05
5B091 BA02 BA03 CA12 CA22 CD15